AD-A235 739

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; Distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): PL-TR-91-2101	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Phillips Lab, Geophysics Directorate; 6b. OFFICE SYMBOL (If applicable): OPE	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): Hanscom AFB Massachusetts 01731-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
61102F	2310	G2	01

11. TITLE (Include Security Classification)
DISCUSSION PAPER - Are Sunspot-Weather Correlations Real?

12. PERSONAL AUTHOR(S)
E.M. Dewan, R. Shapiro

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Reprint	FROM ____ TO ____	1991 May 6	4

16. SUPPLEMENTARY NOTATION
Reprinted from Journal of Atmospheric and Terrestrial Physics, Vol.53, No. 1/2, pp 171-174, 1991

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Sun spots, Solar-weather correlation, "Nonsense-Correlations", Statistical pitfalls, QBO Oscillations, Time series

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We show here that the 17 points of data (which seemed to show a correlation between the sunspot cycle and the temperature of the Northern Hemisphere) are too few in number to permit any demonstration of statistical significance. Labitzke and Van Loon did not realize this; therefore they mistakenly believed that their results could not be explained by pure chance. The present paper exposes the flaws in their reasoning and re-opens the debate on how to interpret their remarkable data.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: E.M. Dewan	22b. TELEPHONE (Include Area Code): (617) 377-4401; 22c. OFFICE SYMBOL: OPE

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Journal of Atmospheric and Terrestrial Physics, Vol. 53, No. 1/2, pp. 171–174, 1991.
Printed in Great Britain.

0021-9169/91 $3.00 + .00
Pergamon Press plc

DISCUSSION PAPER

Are sunspot–weather correlations real?

E. M. DEWAN* and R. SHAPIRO†

*GL/OPE, Hanscomb AFB, MA 01731, U.S.A.; †30 Wayne Rd, Needham, MA 02194, U.S.A.

(*Received in final form* 13 *July* 1990)

A discussion of 'Associations between the 11-year solar cycle, the QBO and the atmosphere. Part I' [LABITZKE K. and VAN LOON H. (1988), *J. atmos. terr. Phys.* **50**, 197].

LABITZKE and VAN LOON (1988), in the above paper, have published data that are striking and suggestive. Their work has received much attention and has inspired further research. The purpose of the present note is to reconsider the question of whether their results could be accounted for by chance alone. Our discussion which is intended to be constructive will be mainly limited to their paper.

The problem can be divided into three sub-questions. (a) What is the probability that a particular 'QBO transformation' of the stratospheric North Pole temperature time-series reveals an 11-yr solar cycle? (b) What is the probability that three cycles of a given geophysical variable would exhibit the same period as the solar cycle? And finally (c), what is the 'statistical significance' of the 0.76 correlation found by them in view of the fact that the samples of the two time-series (temperature and solar-flux) are very sinusoidal in shape? Most of our discussion is in regard to this last question because we wish to show that the statistical test employed by Labitzke and van Loon is inconclusive. Thus the main point of our paper will not be merely to urge caution. Instead our main point will be to show that, in reality, Labitzke and van Loon did not do anything which validly establishes statistical significance.

The first question, (a), raises the issue 'how many other transformations were tried first'? It would make a difference if the QBO filter were the very first one tried or just the last of very many. It will certainly be more impressive if their transformation continues to succeed equally well for the next several cycles without any subsequent 'rule changes'. This issue, among others, has recently received attention by BALDWIN and DUNKERTON (1989) and we recommend their paper to the interested reader. Briefly, we do not see how a rigorous test can be made except by now repeating *exactly* what they have done to the *next* several cycles.

With only 32 data points, any *a posteriori* association must be considered as tentative. The fact that the association was not predicted *before* examining the data makes it impossible to assign any realistic significance level to the association. The tacit assumption in the Monte-Carlo test is that 'x' successes obtainable from 'y' independent trials (i.e. particular transformations applied) determines the chance probability of a success. With a small and *well studied* sample of data, the fact that the association is *a posteriori* implies of necessity that y in this case does not equal unity. A referee has written that 'certainly there may have been subconscious filtering prior to actually writing out the series—no one can really say'. This is our point. What value can one assign to 'y'? Is it 3, 10, or significantly more than 10? Hadamard's book '*The Psychology of Invention in the Mathematical Field*' (HADAMARD, 1945) suggests the last possibility; but, there is no actual way to determine the answer. With this in mind then, we are faced with the fact that the levels of significance are based on the assumption that 'x' and 'y' equal unity. We do not claim to have the value to assign to 'y'. What we claim is that Labitzke and van Loon are also unable to justify a value for 'y' and, in particular, cannot assume that 'y' = unity. Thus we must urge caution in accepting the association at face value.

Another point to make is that even a totally valid Monte-Carlo experiment of this type would not exclusively indicate that the cause must be of solar origin. An alternate physical possibility that cannot be ruled out is, for example, that the nonlinear dynamics of the Earth's atmosphere have caused a long-period, self-sustained oscillation of approximately the same length as that of the Sun's cycle. Later on we shall mention yet another possible interpretation due to TEITELBAUM and BAUER (1990).

The second question is: 'What is the probability

that a geophysical variable exhibits a 'solar cycle' period?" TREFIL (1987) devotes part of a chapter of his book to this question. He listed a number of past observations of correlations between the solar cycle and geophysical variables ranging from air pressure in India (low at sunspot maximum) to the depth of Lake Victoria in Africa. Of these several examples, all have subsequently failed. Trefil went on to show that the lengths of women's skirts (relative to their height) varied between 1926 and 1980 with a solar cycle period. [van Loon (private conversation) has shown one of us (E. D.) an amazingly high correlation between sunspot numbers and the number of Republicans in the U.S. Senate between 1965 and 1985.] Trefil concluded that in view of the large number of geophysical fluctuations which one can try to correlate with sunspots it is more or less guaranteed that one is bound to find that some of them do correlate. The reader may find it useful to consult his book for his numerical arguments; however, he concludes that he would not take seriously any correlation with less than eight cycles.

We now arrive at the third question, i.e. of the assignments of a level of 'statistical significance' to the correlation coefficient (of value 0.76) found by Labitzke and van Loon. They used the following procedure. First they quoted PANOFSKY and BRIER (1963). They gave as a test for significance at the 95% level

$$r_{95} = \frac{1.96}{\sqrt{n_e - 2}} \tag{1}$$

where r_{95} is the correlation coefficient lower bound for the 95% level of significance and n_e is the effective number of independent samples. The latter number can be derived via a procedure which was first discussed by DAVIS (1976). But first it should be noted that the idea is to take into account the non-zero auto-correlations in the two series here considered where such a need is pointed out, for example, in JENKINS and WATTS (1968). Historically this problem was first pointed out by YULE (1926). He showed that the ordinary tests for the statistical significance of correlation coefficients assume that the two datasets being compared resemble *purely random noise*. In other words, each successive point is independent of preceding points. For example, sine waves and other smooth functions violate this requirement.

The procedure of DAVIS (1976) consists of the following. First one calculates the normalized auto-correlations of each of the two series

$$C_T(\tau) = \int \frac{\langle T(t)T(t+\tau)\rangle \,dt}{\langle T^2(t)\rangle}$$

$$C_S(\tau) = \int \frac{\langle S(t)S(t+\tau)\rangle \,dt}{\langle S^2(t)\rangle}$$

where $T(t)$ and $S(t)$ correspond to the temperature and solar flux cycles and the brackets indicate averages Since these are sampled series, one defines Δt as the sampling interval. He then calculated

$$\tau_e \equiv \sum_{i=-k}^{k} C_T(i\Delta t)C_S(i\Delta t)\Delta t. \tag{2}$$

From this one can define

$$n_e \equiv \frac{N\Delta t}{\tau_e}$$

where N is the total number of points.

Next, after doing this, Labitzke and van Loon apply the above to their two series in their fig. 1a. [Solar 10.7-cm flux and 30 mbar North Pole [(January and February)/2] temperature ($N = 32$, $r = (0.14)$).]

They find that N must be reduced from 32 to 23. The resulting $r_{95} = 0.43$, and since the measured value is 0.14, they say, correctly, that there is no statistical significance. After treating other questions they again return to matters of correlation and significance i.e. they consider fig. 1b which is the 'QBO' transformed version of their temperature data. Here we see two nearly, perfectly in-phase, sinusoidal oscillations, one superposed on the other, with a duration of about three periods (1956–1987). The correlation coefficient here is $r = 0.76$; and, at this point more than ever, one needs to take into account the effects of the auto-correlations. Labitzke and van Loon then state: 'If we assume that each of the winter temperatures are independent (i.e. $n_e = 17$) this correlation is significant above the 99% level. Even if one reduces the number of independent samples by 50% (i.e. $n_e = 9$) it is significant at the 95% level.' Unfortunately, as will be discussed below, the number of degrees in the data under consideration is close to four. In any case it should be noted that they gave as a reason why they did not calculate n_e from their data the following statement: 'because the sampling interval varies'. Perhaps a better reason is that, in the use of (2) to calculate n_e, the probable errors in C_T and C_S are too large ($C_T = 0.4$ cannot be distinguished from zero for $N = 17$); this worsens for each succeeding value of the auto-correlations.

With each succeeding lag of the auto-correlation function the sample size decreases and the error bounds increase. Therefore, little confidence can be placed on any estimate of τ_e from equation (2) if the total sample size (N) is only 17. However, there is a more serious problem involving the lack of station-

arity of the time series from which the auto-correlation function is obtained.

With each succeeding value of the auto-correlation function, the sample size decreases by 1. Consequently, with a small basic sample (17 in this case), end effects become important. That is, the sample of 16 from which lag 1 is obtained is of necessity different from and inconsistent with the sample of 12 from which lag 5 is obtained, and each of these samples are in turn inconsistent with the samples from which the other lag correlations are obtained. The inhomogeneity among the auto-correlation coefficients is exacerbated with increasing lag and makes it impossible to obtain a valid estimate of n_e.

It is nevertheless appropriate here to discuss further the estimated number of degrees of freedom in the solar activity and temperature sinusoidal time-series that was mentioned earlier. If these series were in fact two perfect sine waves, then there would be four degrees of freedom in view of the fact that only the phase and period of each sine wave matters in the present context. In other words, since each sine wave has two degrees of freedom, the total must be four. Of course the actual time-series are both not perfect sine waves and this then raises the question of whether or not there are more degrees of freedom due to this fact. In order to answer this question the reader should imagine what would happen to the two time-series if a least-squares fitted sine wave were subtracted out of each of them. Would there remain any compelling correlation after this procedure is carried out? It appears to be obvious that the answer is 'no'. To the extent that this 'no' answer is correct, the estimate of four degrees of freedom is also correct; but, a skeptical reader who needs further proof about this may wish to consult PIERCE (1977) for more discussion on this approach. If one then assumes that the correlation, $r = 0.76$, and the number of degrees of freedom, d.f. $= 4$, then the value of 't' in the 't-test' for level of significance of r (MORONEY, 1951), is given by

$$t = (r\sqrt{\text{d.f.}}/\sqrt{1-r^2}) = 2.34.$$

But in order to obtain a 5% level of significance, t must be at least as large as 2.78. Therefore the correlation is not significant.

It should now thus be clear that Labitzke and van Loon have not yet shown valid quantitative evidence for a statistically significant solar–weather correlation. Is there another, more valid test in this regard for one to try? The answer is 'perhaps' and it is described in GOTTMAN (1981).

In essence this later technique (called the Gottman–Ringland procedure) compares two time-series models. Let S_t and T_t represent the solar flux and North Pole temperatures as a function of time t. The two models are

$$T_t = \sum_{i=1}^{B} b_i T_{t-i} + e_t \tag{3}$$

and

$$T_t = \sum_{i=1}^{B} b_i T_{t-i} + e_t + \boxed{\sum_{i=1}^{C} D_i S_{t-i}}$$

$$S_t = \sum_{i=1}^{A} a_i S_{t-i} + n_t \tag{4}$$

where n_t and e_t represent random noise and the term in the box includes a casual influence. Significance is determined by means of a likelihood ratio test to see if model (4) does a significantly better job of prediction than (3).

This test has not yet been applied to the data of fig. 1b of Labitzke and van Loon. However, and this is most unfortunate, the Gottman–Ringland procedure again must rely on auto-correlations which, as we already pointed out above, cannot be calculated with the requisite confidence when one has merely 17 points at one's disposal.

The reader may wonder why we mention this procedure in view of the fact that it is not valid to apply it to so few data. Our purpose has been to show what would *actually* be needed to provide statistical evidence of the cause and effect relationship that has been claimed.

The issue of the effective sample size (or the equivalent number of independent data points) in data samples with auto-correlation in time (or space) has been extensively examined by THIEBAUX and ZWIERS (1984). Using a Monte-Carlo approach with sample sizes ranging from 30 to 240, with two different data generating models and a number of values of the lag 1 auto-correlation, they compare seven different methods of estimating effective sample size (ESS). They conclude that ESS is difficult to estimate reliably and that, without knowledge of the power spectrum of the observed process, ESS cannot be estimated reliably without a large data sample. Even with the largest data samples ($N = 240$) there is considerable variability in the estimates of ESS among the seven methods, although one or two methods succeed in approximating the 'true' (model generated) values of ESS. However, with $N = 30$ (almost twice the size of the Labitzke and van Loon sample) there is no confidence in any of the estimates of ESS.

It should be clear now that, while the data of

Labitzke and van Loon are very interesting, suggestive, and even dramatic, these authors have not shown that the relationship could not have occurred by chance alone, notwithstanding the authors' positive Monte-Carlo experiment.

BARNSTON and LIVEZEY (1989) have addressed this question by carrying out a very extensive series of Monte-Carlo experiments with essentially the same data examined by Labitzke and van Loon. They urge 'caution and reserve' in accepting the reality of the relationship in spite of a number of apparently statistically significant relationships between the solar flux and the QBO filtered large-scale upper air pressure and surface temperature fields.

A recent and very thought provoking paper has been published by TEITELBAUM and BAUER (1990) which suggests a possible explanation of the observations of LABITZKE and VAN LOON (1988). They showed that the application of the QBO transformation to the data could *itself* be the cause of the solar-cycle period that was observed in the temperature data. They reason that the QBO possibly modulates the stratospheric temperature so that the latter has a cycle of about 27.7 months. The sampling procedures involved, however, lead to an approximately 24-months sampling interval. Thus there would be a phase shift observed of 3.7 months for each data point leading to a complete cycle of shift in about 15 yr. This 'stroboscopic' or 'alias' effect would generate a cycle with period indistinguishable from a solar period in the sample length of 32 yr presently available. Further data may either confirm or invalidate this stroboscopic effect. We feel that this possibility can only further emphasize the present need for caution.

In conclusion, we feel that it is not possible to demonstrate reality with the available, small, non-independent data samples. In this connection we think that further attempts at establishing statistical significance of a result based upon such a small sample would be unproductive. We believe that a demonstration (or reality) must be based on independent data in which the expected relationship is maintained, or on a solid physical theory which itself predicts the postulated relationships. On the other hand, it is hoped that our remarks will bring about a further healthy debate and hence more clarification on this highly interesting subject.

Acknowledgements—The authors thank Dr Richard Picard for bringing to their attention the book by Trefil. Also James Brown is thanked for comments. We also acknowledge the fact that this research was sponsored in part by AFOSR under project MAPSTAR, 2310G2.

REFERENCES

BALDWIN M. P. and DUNKERTON T. J. 1989 *Geophys. Res. Lett.* **16**, 863.
BARNSTON A. G. and LIVEZEY R. E. 1989 *J. Climate* **3**, 1295.
DAVIS R. E. 1976 *J. phys. Oceanogr.* **6**, 249.
GOTTMAN J. M. 1981 *Time-series Analysis* (p. 321). Cambridge University Press, Cambridge.
HADAMARD J. 1945 *The Psychology of Invention in the Mathematical Field.* Dover Press, New York.
JENKINS G. M. and WATTS D. G. 1968 *Spectral Analysis and its Applications* (p. 337). Holden-Day, San Francisco.
LABITZKE K. and VAN LOON H. 1988 *J. atmos. terr. Phys.* **50**, 197.
MORONEY M. J. 1951 *Facts from Figures* (p. 311). Penguin Books, Baltimore.
PANOFSKY H. A. and BRIER G. W. 1963 *Some Applications of Statistics to Meteorology* (p. 92). Pennsylvania State University, University Park.
PIERCE D. A. 1977 *J. Am. Stat. Assn* **73**, 11.
TEITELBAUM H. and BAUER P. 1990 *Ann. Geophys.* **8**, 239.
THIEBAUX H. and ZWIERS F. 1984 *J. Climate appl. Meteor.* **23**, 800.
TREFIL J. 1987 *Meditations at Sunset* (Chapter 3). Charles Scribner's Sons, New York.
YULE G. U. 1926 *J. R. Stat. Soc.* **89**, 1.